UDC 551.509.314:551.509.54

Use of Model Output Statistics for Predicting Ceiling Height

JOSEPH R. BOCCHIERI and HARRY R. GLAHN—*Techniques Development Laboratory, National Weather Service, NOAA, Silver Spring, Md.*

ABSTRACT—The model output statistics (MOS) technique consists of determining a statistical relationship between the forecast output of numerical prediction models and a predictand. This paper presents some results obtained in applying the MOS technique to the prediction of ceiling height by means of screening regression.

Data from 3 winter seasons and 95 eastern U.S. stations are combined in a generalized operator approach to develop multiple regression equations. The potential predictors subjected to screening include surface variables observed at 0700 GMT and forecast output from both the National Meteorological Center's primitive-equation model and the Techniques Development Laboratory's subsynoptic advection model. Prediction equations are developed for 5-, 11-, and 17-hr forecast projections representing ceiling height forecasts valid at 1200, 1800, and 2400 GMT, respectively.

Ceiling height is treated both as a categorized and as a continuous predictand. Where ceiling height is categorized, the regression estimation of event probabilities (REEP) screening technique is used to develop probability forecast equations. Where ceiling height is treated as a continuous variable, specific ceiling height forecast equations are developed by ordinary screening regression.

The independent sample used for testing consists of data for 20 stations in the eastern United States from the winter of 1970–71. Several verification scores, including the Brier *P*-score, Allen utility score, Heidke skill score, and percent correct, are presented. The verification results indicate that forecasts from the REEP equations are generally better than those from the equations that produce specific heights. Also, the REEP forecasts are better than persistence and climatology.

1. INTRODUCTION

The purpose of this study was to apply the model output statistics (MOS) technique to the prediction of ceiling height. The MOS technique consists of determining a statistical relationship between the forecast output of numerical prediction models and a predictand. In this study, the statistical relationship was determined by screening regression.

According to Roberts (1967), the aviation terminal forecast is the most difficult and demanding of all those forecasts that the National Weather Service (NWS) issues (e.g., public, marine, agricultural, and fire weather). This difficulty is due not only to the high degree of precision and time resolution that these forecasts require but also to the complexity of the atmospheric processes that produce changes in terminal weather elements such as ceiling and visibility. Many of these atmospheric processes operate on the microscale or the mesoscale. Basically, then, the problem involves a microscale or mesoscale phenomenon, whereas the existing forecast and observing system is geared to the synoptic scale.

Some years ago, extensive work was carried out at the Travelers Research Center, Hartford, Conn., in the area of objective terminal forecast systems. One result of this work was the development of a screening technique known as regression estimation of event probabilities (REEP) (Miller 1964). Enger et al. (1964) used the REEP technique for the 2- to 7-hr prediction of ceiling and visibility and found that it produced forecasts of approximately the same accuracy as both the best objective techniques then available and manually prepared subjective forecasts.

In recent years, NWS has been conducting research and development in terminal weather forecasting for the Federal Aviation Administration (FAA). An operational test (Allen 1969) of an objective forecast system based on the REEP technique was conducted by NWS during the 7-mo period, September 1966 through March 1967. The purpose of the test was to examine the value of REEP forecasts as guidance for aviation forecasters. The results indicated that the REEP forecasts were of some value for this purpose. However, the value was found to be small and irregular because these particular equations did not provide the forecaster with high probabilities of forthcoming low ceiling and visibility conditions. The experiment also demonstrated the feasibility of computing and distributing such automated forecasts from a central computer at the National Meteorological Center (NMC).

The REEP equations that produced the forecasts for this operational test used as predictors simple variables observed at a network of stations surrounding the predictand station. Because of the difficulties encountered in processing data from a network of stations and in defining predictors from that network, it seemed desirable to see how much accuracy would be lost by using predictors from only the predictand station (single-station approach). Allen (1970*a*) compared network equations with single-station equations and found that the network equations were better, but the increase in accuracy was insufficient to justify the much greater cost of deriving such equations.

This study describes an effort to apply forecast output from numerical prediction models to the prediction of ceiling height by the MOS approach. The MOS approach has produced useful results in the prediction of precipitation probability (Glahn and Lowry 1969), surface wind (Glahn 1970, Barrientos 1970), maximum temperature (Annett et al. 1972), and conditional probability of frozen precipitation (Glahn and Lowry 1972). The two numerical models used in this study are run operationally at NMC—the primitive equation (PE) model (Shuman and Hovermale 1968) and the subsynoptic advection model (SAM) (Glahn et al. 1969). Ceiling height was treated both as a categorized and a continuous predictand. Where ceiling height was categorized, the REEP screening technique was used to develop probability forecast equations. Where ceiling height was treated as a continuous variable, specific ceiling height forecast equations were developed by ordinary screening regression (Miller 1958).

2. DATA SAMPLE AND DEFINITION OF PREDICTORS

The MOS technique requires that a sample of forecasts from a numerical model (or models) be collected for analysis. Such a sample of predictor data from the SAM and PE models, as well as predictand data, was available from the Techniques Development Laboratory's (TDL) SAM Project (Glahn et al. 1969). The developmental data sample consisted of 454 days from the 3 winter seasons (October–March) of 1967–68 through 1969–70. Since the developmental sample was small, the generalized operator approach was used. This approach consists of pooling the data for all stations in the development of a forecast equation. The same equation is then considered applicable to all stations. In this instance, a 40,000-case sample resulted from combining data from 95 stations in the eastern United States. Applications of the generalized operator approach are contained in the works of Harris et al. (1963) and Russo et al. (1966).

Ceiling prediction equations were developed for 5-, 11-, and 17-hr forecast projections, representing ceiling height forecasts valid at 1200, 1800, and 2400 GMT, respectively (the "today" period). The forecasts used from the PE model (0000 GMT initial time) and SAM (0700 GMT initial time) were also valid during this same period and were interpolated from gridpoints to stations. The predictors subjected to screening regression included not only PE and SAM forecast output but also "initial" hourly observations made at 0700 GMT. These initial data include weather, clouds, ceiling, visibility, temperature, dew point, sea-level pressure, and wind components observed at the surface.

Forecast output from SAM included saturation deficit, sea-level pressure, 1000-mb geostrophic wind components, and indicators of terrain-induced vertical velocities. The

TABLE 1.—*Potential predictors for the development of ceiling prediction equations*

0700 GMT observed surface data
Temperature (°F)
Dew point (°F)
Weather (4 classes)
Total cloud cover (5 classes)
U wind component (kt)
V wind component (kt)
Wind speed (kt)
Ceiling height (ft)
Visibility (mi)
SAM forecast output valid at 1200, 1800, and 2400 GMT
Sea-level pressure (mb)
U component of 1000-mb geostrophic wind (kt)
V component of 1000-mb geostrophic wind (kt)
Upslope component of 1000-mb geostrophic wind (mb/s$\times 16\times 10^{-6}$)
Saturation deficit (m)
Cumulative 3-hourly saturation deficit (m)
PE forecast output valid at 1200, 1800, and 2400 GMT
Mean relative humidity, sfc.–400 mb (percent)
Vertical velocity at 1000 mb (10^{-3} mb/s)
Vertical velocity at 850 mb (10^{-3} mb/s)
Vertical velocity at 500 mb (10^{-3} mb/s)

saturation deficit is a term defined by Younkin et al. (1965) as

$$S_d = h_5 - S_T \tag{1}$$

where h_5 is the 1000- to 500-mb thickness and S_T is the saturation thickness. For SAM, the saturation thickness is defined as that thickness between 1000 and 500 mb for which precipitation will occur for a given amount of moisture between those levels.

Forecast output from the PE model included the mean relative humidity in the column from the surface to about 400 mb and vertical velocities at 1000, 850, and 500 mb.

Table 1 summarizes the predictors subjected to screening regression in this study.

3. DEVELOPMENT OF PROBABILITY FORECAST EQUATIONS

As discussed by Keaty and Allen (1970), REEP screening is a statistical technique for selecting a subset of effective predictors from a set of possible predictors to use in multiple linear regression equations. The equations developed give estimates of the probabilities of occurrence for a given set of predictands. Before input to the REEP program, the predictands and predictors are "dummied"; that is, they are converted into dichotomous, binary-type variables. Each dummy variable takes the value of either one or zero for a given occasion depending, respectively, upon whether the variable occurs or does not occur. The predictands in a given set must be mutually exclusive and, collectively, cover all possible values of the predictand variable.

Table 2 gives the limits of the five ceiling height predictands used in this study and the relative frequency of occurrence of each predictand at 1200 GMT based on the developmental data sample. The five categories provide the kind of detail needed for operational decisions and have been designated by the FAA under Interagency Agreement FA67WAI–131. The narrow categories at low values of ceiling have a low frequency of occurrence and are extremely difficult to forecast.

Table 3 shows an example of REEP equations incorporating SAM and PE forecast output and 0700 GMT initial data as predictors. The equations contain the first eight predictors selected by the REEP screening process. Note that a separate regression equation is developed for each predictand category. The predictors are the same for each equation; however, the constants and coefficients vary.

TABLE 2.—*The five predictand categories of ceiling and relative frequency at 1200 GMT based on the developmental data sample (40,130 cases)*

Category	Ceiling (ft)	Relative frequency
1	≤100	0. 015
2	200–400	. 043
3	500–900	. 065
4	1, 000–1, 900	. 084
5	≥2, 000	. 793

TABLE 3.—*Example of REEP forecast equations incorporating numerical model forecasts as predictors. The equations produce forecasts valid at 1800 GMT (11-hr forecast).*

Selected predictors (units given in table 1)	Predictor coefficients ceiling predictand category (ft)				
	≤100	200–400	500–900	1, 000–1, 900	≥2, 000
Constant	0. 004	0. 025	0. 060	0. 057	0. 854
1. SAM cumulative saturation deficit (1800 GMT) ≤15.	. 001	. 026	. 049	. 062	–. 138
2. Observed ceiling (0700 GMT) ≤1,900	. 002	. 017	. 051	. 156	–. 226
3. PE mean relative humidity (1800 GMT) ≤ 90.	–. 004	–. 024	–. 056	–. 033	. 117
4. SAM saturation deficit (1800 GMT) ≤–2.	. 007	. 052	. 095	. 017	–. 171
5. Observed ceiling (0700 GMT) ≤400.	. 023	. 079	. 047	–. 067	–. 082
6. SAM cumulative saturation deficit (1800 GMT) ≤120.	. 000	–. 001	. 019	. 065	–. 082
7. SAM upslope wind component (1800 GMT) ≤–100.	–. 001	–. 009	–. 020	–. 042	. 073
8. PE 850-mb vertical velocity (1800 GMT) ≤ –1.	. 003	. 042	. 058	–. 004	–. 099

To determine the relative usefulness of various predictor types for ceiling forecasting, we developed six sets of REEP equations for each forecast projection. Each of the six sets of REEP equations used different predictor combinations as shown in table 4. For instance, REEP equation set number 1 used only 0700 GMT initial data as predictors (table 1). Equation set number 6 used SAM and PE forecast output in addition to 0700 GMT initial data as predictors. Each set of REEP equations contained 20 predictors except in the case of set 1 where only 17 predictors were included.

Table 4 also shows the reduction of variance, averaged for the five ceiling predictand categories, for each equation set and forecast projection. The following conclusions can be made:

1. Initial 0700 GMT data alone are better than numerical model output at the 5-hr projection. However, each model alone is better than initial data alone at 17 hr. The transition point is near 11 hr.
2. The combination of initial data and numerical model variables is better than either alone for all projections; however, there is little difference at the 5-hr projection.
3. SAM is better for ceiling prediction purposes than the PE model at 5- and 11-hr projections. They are of about equal value at 17 hr.

REEP equation sets 1, 4, and 6 were verified on independent data, and the results will be presented in a later section of this paper.

4. DEVELOPMENT OF SPECIFIC CEILING HEIGHT FORECAST EQUATIONS

In the present terminal forecast (FT) system, NWS forecasters do not give probability forecasts of ceiling height categories to the users. That is, the FT contains a specific ceiling height forecast. Therefore, we decided to develop specific ceiling height forecast equations ("continuous" equations) in addition to the REEP equations described in the previous section.

The dependent data sample, the possible predictors subjected to screening, and the forecast projections were the same as those used in the development of the REEP equations. However, most of the predictors as well as the predictand were left in continuous form rather than being made into binary variables. To treat ceiling as a continuous variable, we assigned a value of 35,000 ft to "unlimited" ceiling observations.

TABLE 4.—*Combination of predictors used in developement of REEP forecast equations and results on dependent data. Reduction of variance (percent) is average for five ceiling predictand categories. Sample had 40,130 cases.*

REEP equation set	Predictor combinations	Reduction of variance forecast projection		
		5 hr	11 hr	17 hr
1	0700 GMT observed surface data	20	10	6
2	SAM output (0700 GMT initial time)	11	12	9
3	PE output (0000 GMT initial time)	8	10	9
4	0700 GMT observed data, SAM output	21	14	10
5	0700 GMT observed data, PE output	20	13	10
6	0700 GMT observed data, SAM, PE output	21	15	11

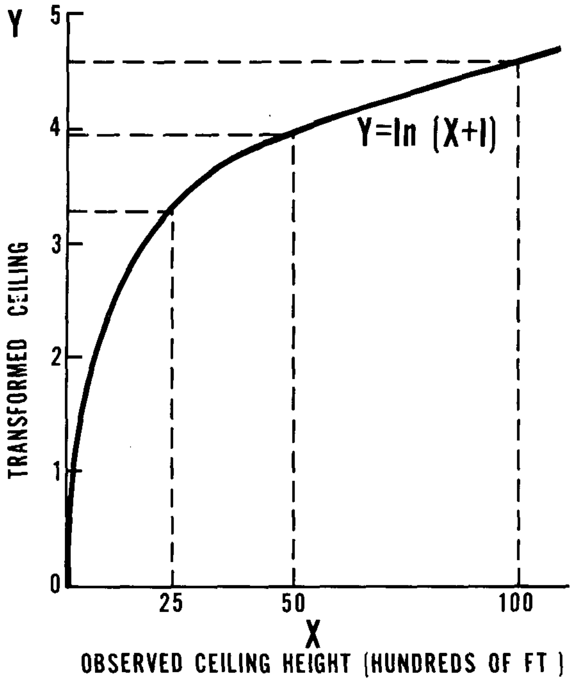

FIGURE 1.—The natural log (ln) transform function applied to ceiling height.

The screening procedure was employed to develop the prediction equations. The method is explained in detail by Miller (1958) and Klein (1965), and an adequate discussion of linear multiple regression is given by Panofsky and Brier (1958). In the development of the continuous forecast equations, an attempt was made to obtain more resolution in the forecast of low rather than high ceiling heights. To do this, we applied a transform function to the observed ceilings; then we used the transformed ceilings as the predictand in the regression analysis. The effect of applying a natural logarithm (ln) transform function to observed ceilings is shown in figure 1. One can see that ceiling height differences at the lower end of the ceiling scale are expanded on the transformed scale in relation to height differences at the upper end of the scale. The transformation is such that more weight will be given to variations within the lower ceiling height ranges in the regression analysis.

To determine an appropriate transform function, we tested the following functions:

1. $Y=X$,
2. $Y=\ln(X+1)$,
3. $Y=\ln(X+0.001)$,
4. $Y=\ln(X+0.0001)$,

and

5. $Y=3X/(X+1)$

where X is the observed ceiling in hundreds of feet, and Y is the transformed ceiling. Function 1, of course, involves no transform at all. Functions 2–4 are natural log transforms with different scaling factors (which are necessary because X can equal zero), and function 5 is a hyperbolic transform.

TABLE 5.—*The categories of ceiling height used for comparative verification of transform functions*

Category	Ceiling height (ft)	Category	Ceiling height (ft)
1	≤100	5	2, 000–4, 900
2	200–400	6	5, 000–9, 900
3	500–900	7	10, 000–19, 900
4	1, 000–1, 900	8	≥20, 000

Five continuous equations were developed for producing 5-hr ceiling forecasts, each incorporating one of the five functions described above. That is, in the development of each regression equation, one of the functions above was used to transform the observed ceiling before the regression analysis was performed. The equations thus developed produced forecasts of a transformed ceiling.

To compare the performance of the five transform functions, we evaluated each of the five forecast equations for 20 eastern U.S. stations and 3 winter seasons (1967–68 through 1969–70). This verification sample (8,661 cases) was a subset of the developmental data sample. Two verification scores, the percent correct and Heidke skill score [defined in the app. by eq (4) and (6), respectively] were computed from a contingency table that utilized the eight ceiling categories defined in table 5. The forecasts were converted back to actual ceiling heights before being entered into the contingency table. Table 6 shows the results of the verification. The following conclusions may be drawn:

1. In terms of percent correct, the equation using the $Y=X$ transform gave the best results for ceiling categories greater than 2,000 ft, but the poorest results for the lower, more important ceiling categories (<2,000 ft).
2. In terms of percent correct, the equations using the natural log transforms gave the best results for the lower ceiling categories and were better than the hyperbolic transform, $Y=3X/(X+1)$, for the higher ceiling categories.

On the basis of the preceding results, we decided that the natural log transform $Y=\ln(X+0.001)$ was the most desirable of those tested for specific ceiling height forecasting. Because of the rather close relationship of the predictand to the initial ceiling, the transform was also applied to the 0700 GMT ceiling predictor. A sample continuous forecast equation is shown in table 7.

5. INDEPENDENT DATA EXPERIMENTS AND RESULTS

The independent data sample consisted of data pooled for 20 eastern U.S. stations from the winter season of 1970–71 (about 2,500 cases). The 20 stations used are listed in table 8; the "WBAN" station numbering system is that used by the National Climatic Center, Asheville, N.C.

Three sets of REEP equations and one continuous equation were verified on independent data. The first set of REEP equations uses only 0700 GMT initial data as

TABLE 6.—*Verification scores for 5-hr specific ceiling height forecasts made with various transform functions. Dependent data pooled for 20 eastern U.S. stations (8,661 cases).*

Transform functions used in regression equations	Percent correct (all categories)	Percent correct (<2,000 ft)	Percent correct (≥2,000 ft)	Skill score
1. $Y = X$	47. 2	18. 9	48. 2	0. 27
2. $Y = \ln(X+1)$	43. 0	29. 2	45. 2	. 28
3. $Y = \ln(X+0.001)$	41. 0	30. 7	42. 8	. 26
4. $Y = \ln(X+0.0001)$	40. 2	30. 7	41. 9	. 25
5. $Y = 3X/(X+1)$	30. 1	25. 1	32. 7	. 16

TABLE 7.—*Sample specific ceiling height forecast equation for an 11-hr forecast valid at 1800 GMT. The ceiling predictor and predictand are transformed by the function $Y=\ln(X+0.001)$.*

Selected predictors (units given in table 1)	Coefficient	Cumulative reduction of variance
Constant	3. 28	
1. SAM cumulative saturation deficit (1800 GMT)	0. 004	0. 358
2. Observed ceiling (0700 GMT)	. 238	. 421
3. PE 850-mb vertical velocity (1800 GMT)	. 168	. 441
4. SAM upslope wind component (1800 GMT)	—. 001	. 449
5. PE mean relative humidity (1800 GMT)	—. 008	. 455

TABLE 8.—*The 20 stations used for verification of ceiling height forecast equations on independent data*

Station	WBAN No.	Station	WBAN No.
Albany, N.Y.	14735	Knoxville, Tenn.	13897
Atlanta, Ga.	13874	Louisville, Ky.	93821
Baltimore, Md.	93721	Nashville, Tenn.	13897
Birmingham, Ala.	13876	New Orleans, La.	12916
Boston, Mass.	14739	Pittsburgh, Pa.	94823
Buffalo, N.Y.	14733	Raleigh-Durham, N.C.	13722
Chicago Midway Airport, Ill.	14819	Savannah, Ga.	03822
Cincinnati, Ohio	93814	St. Louis, Mo.	13994
Cleveland, Ohio	14820	Tallahassee, Fla.	93805
Kennedy Airport, N.Y.	94789	Washington, D.C.	13743

predictors. The second set uses 0700 GMT initial data and SAM forecast output as predictors. The third REEP set and the continous equation use 0700 GMT initial data as well as SAM and PE forecast output as predictors. All of these equations were verified for 5-, 11-, and 17-hr forecast projections representing ceiling height forecasts valid at 1200, 1800, and 2400 GMT, respectively. Several verification scores were computed, including the Brier *P*-score, Allen utility score, percent correct, and Heidke skill score, all of which are defined in the appendix.

The independent data verification had the following purposes:

1. To determine the number of predictors that should be included in the REEP and continuous forecast equations.
2. To compare forecasts from the three REEP equation sets and determine if the use of numerical model output as predictors improves upon the use of 0700 GMT initial data only.
3. To compare forecasts from the best REEP equation set determined previously with those from the continuous forecast equations.
4. To compare the objective forecasts with persistence and climatology.

To determine the number of predictors that should be included in the REEP forecast equations for each forecast projection, we computed all of the verification scores for each of the equations containing 2, 4, 6, . . . , 18, and 20 predictors and compared them. In the case of the continuous method, all the scores except the Brier *P*-score were computed for each of the equations containing 1, 2, 3, . . ., 14, and 15 terms.

Figures 2 and 3 show the results of computing the Brier *P*-score and the Allen utility score for the 5-, 11-, and 17-hr REEP ceiling forecast equations incorporating SAM and PE forecast output and 0700 GMT initial data as predictors. Inspection of these figures indicates a generally steady improvement (decrease) in the Brier *P*-score and improvement (increase) in the Allen utility score as the number of predictors in the REEP equations is increased to about 12–14, with the scores then remaining about constant out to 20 predictors. Figure 4 shows the results of computing the Heidke skill score for the 5-, 11-, and 17-hr continuous equations. Figure 4A depicts an increase (improvement) in the Heidke skill score at 5 hr as the number of predictors is increased to four, with a general decrease in the score as more predictors are added. Figures 4B and 4C indicate a general increase in the Heidke skill score at 11 and 17 hr as the number of predictors is increased to about 10, with the score remaining about the same thereafter.

Similar graphs were drawn for all verification scores and all forecast projections for both the REEP and continuous forecast systems, and generally about the same results were obtained. These graphs indicate that the REEP equations should contain 14 to 20 predictors, and the continuous equations should include about 10 predictors.

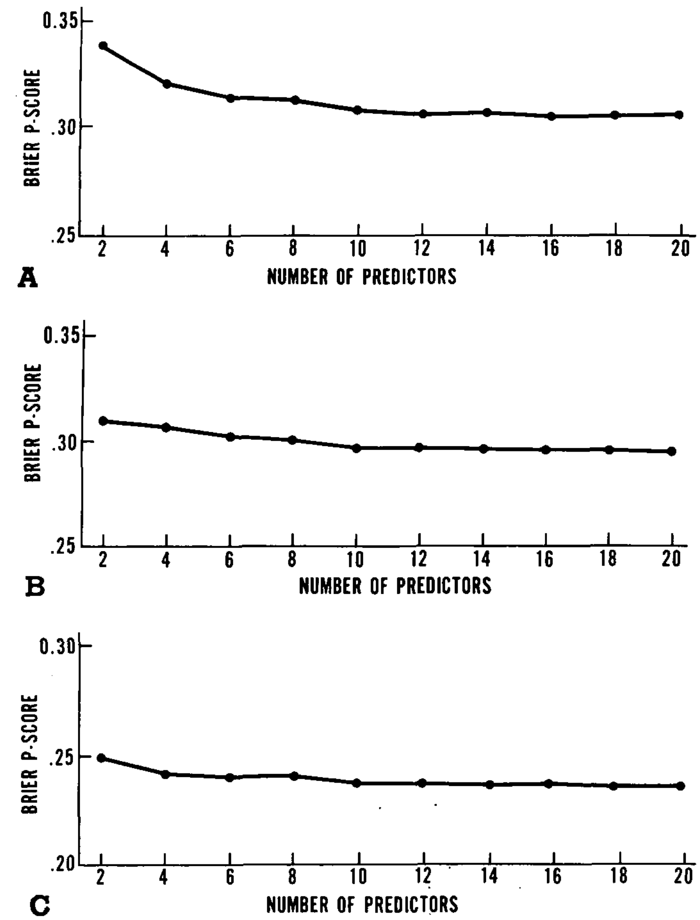

FIGURE 2.—The Brier *P*-score computed for the (A) 5-hr (2,519 cases), (B) 11-hr (2,527 cases), and (C) 17-hr (2,519 cases) REEP forecast equation sets.

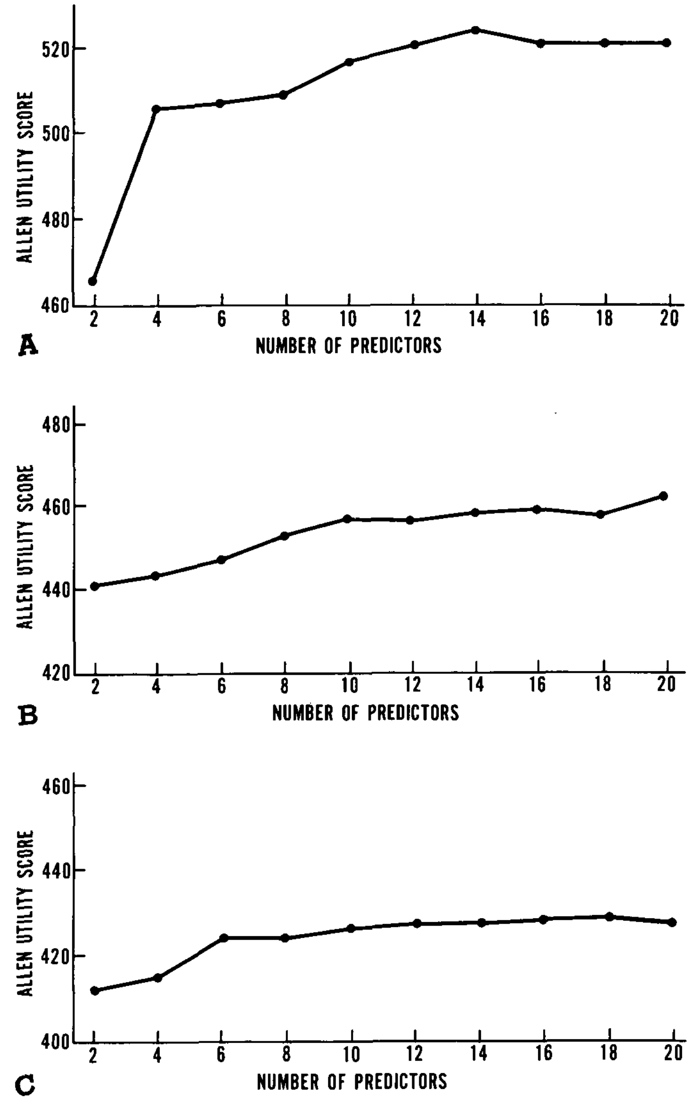

FIGURE 3.—The Allen utility score computed for the (A) 5-hr (2, 519 cases), (B) 11-hr (2,527 cases), and (C) 17-hr (2,519 cases) REEP forecast equation sets.

TABLE 9.—*Verification scores for three sets of REEP forecast equations for 5-, 11-, and 17-hr ceiling forecasts. Independent data are pooled for 20 eastern U.S. stations.*

REEP equation set	Brier *P*-score	Allen utility score	Heidke* skill score	Percent† correct (<2,000 ft)	Percent correct (all categories)
5-hr forecast (2,519 cases)					
0700 GMT	0.312	515	0.349‡	37.9	77.6
0700 GMT, SAM	.304‡	522‡	.340	41.7‡	78.0‡
0700 GMT, SAM, PE	.304‡	521	.337	41.6	78.0‡
11-hr forecast (2,527 cases)					
0700 GMT	.314	440	.054	28.9	78.6
0700 GMT, SAM	.299	452	.134‡	34.2‡	79.0‡
0700 GMT, SAM, PE	.296‡	461‡	.119	31.2	79.0‡
17-hr forecast (2,519 cases)					
0700 GMT	.258	394	.011	16.6 (6)	84.2
0700 GMT, SAM	.242	416	.028	41.6 (12)	84.2
0700 GMT, SAM, PE	.235‡	428‡	.043‡	56.3 (16)‡	84.4‡

*The skill score was not maximized for this table; it was computed from the contingency table that maximized the number correct. The Allen utility score was maximized. (See app. for methods used to maximize various verification scores.)

†Percent correct for the first four ceiling categories only (table 2). For the 17-hr projection, the numbers in parentheses indicate the number of forecasts <2,000 ft.

‡The best equation set for each score and forecast projection.

Table 9 shows a comparison of three REEP ceiling forecast equation sets for all verification scores and forecast projections. The results indicate that the REEP equation set using 0700 GMT initial data and numerical model output as predictors did better than the REEP equation set using only 0700 GMT initial data as predictors for all projections and all scores except the Heidke skill score for the 5-hr projection. Note also that the REEP equation set using SAM and PE forecast output and 0700 GMT initial data as predictors becomes the best of the three sets as the forecast projection increases from 5 to 17 hr. On the basis of these results, we decided that the latter was the most appropriate probability forecast equation for ceiling height.

Table 10 shows the verification results on independent data for 5-, 11-, and 17-hr ceiling height forecasts produced by four forecast systems that are defined as follows:

1. *REEP system.* This equation set uses 20 predictors (SAM and PE forecast output and 0700 GMT initial data) and produces probability forecasts for five ceiling height categories (table 2). The verification scores for this forecast system were not computed from the same set of categorical forecasts; that is, score definition and the probability forecasts were used to choose the category that would, theoretically, maximize *that particular score.* Methods for maximizing the various scores are presented in the appendix.

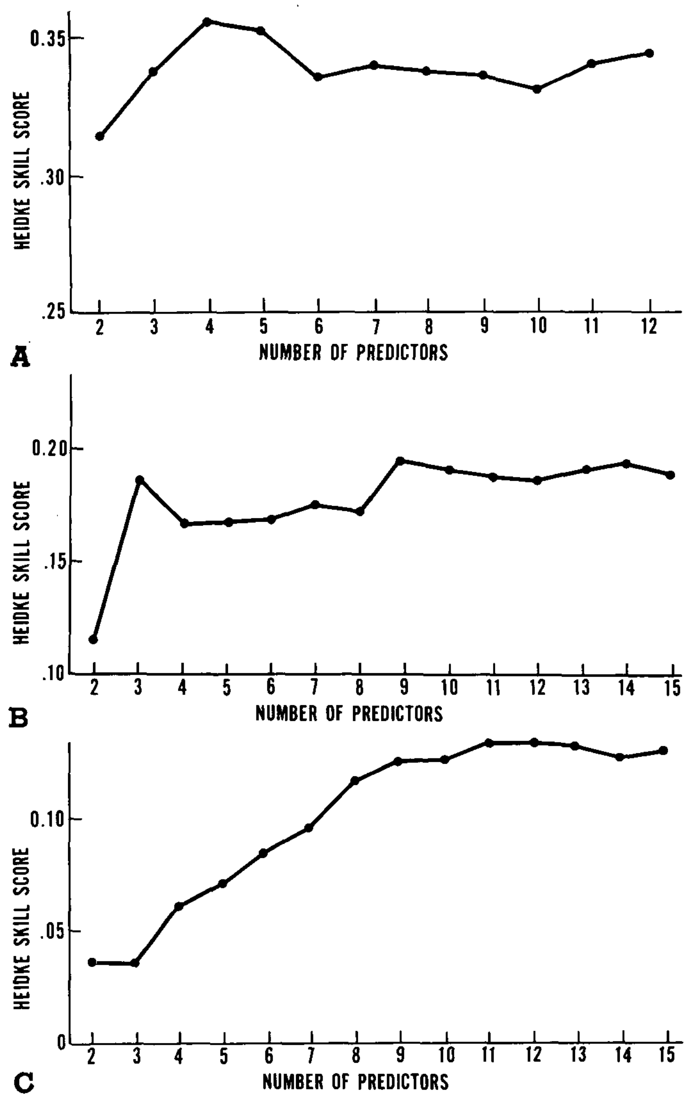

FIGURE 4.—The Heidke skill score computed for the (A) 5-hr, (B) 11-hr, and (C) 17-hr continuous forecast equation. Independent data sample size was 2,519 cases.

2. *Continuous system.* This equation uses 10 predictors (0700 GMT initial data, SAM, and PE forecast output) and produces specific ceiling height forecasts. The continuous equation incorporates the natural log function $Y=\ln(X+0.001)$, which was determined to be the most desirable of several transformations tested for obtaining more resolution in forecasts of low ceilings.

3. *Persistence.* This method uses the ceiling height category observed at 0700 GMT as a forecast for 1200, 1800, and 2400 GMT.

4. *Climatology.* This system gives a constant probability forecast equal to the relative frequency of occurrence of each of the five ceiling height categories in table 2 for each station on the dependent data. This system always produced a forecast of ceiling height category 5.

The following conclusions were made from the verification results shown in table 10.

1. The REEP forecast system is better than the continuous forecast system.

2. The percent-correct score shows REEP to be only slightly better than persistence at the 5-hr forecast projection. However, REEP shows definite improvement over persistence at the 11- and 17-hr forecast projections. The Allen utility score and Heidke skill score show definite improvement of REEP over persistence at all forecast projections.

TABLE 10.—*Verification scores for four ceiling height forecast systems for 5-, 11-, and 17-hr forecast projections. Independent data pooled for 20 eastern U.S. stations (2,519 cases)*

Forecast system	Brier P-score	Allen utility score	Heidke skill score	Percent correct (all categories)	Percent* correct
		5-hr forecasts valid at 1200 GMT			
REEP	0.304†	521†	0.421†	78.0†	45.2 (230)†
Continuous	–	415	.342	76.3	27.3 (175)
Persistence	–	476	.416	77.6	40.1 (309)
Climatology	.417	365	–	75.0	–
		11-hr forecasts valid at 1800 GMT			
REEP	.296†	459†	.300†	78.9†	32.6 (46)
Continuous	–	359	.190	78.7	38.1 (21)†
Persistence	–	397	.243	73.1	14.6 (309)
Climatology	.359	366	–	78.6	–
		17-hr forecasts valid at 2400 GMT			
REEP	.235†	428†	.284†	84.4†	56.3 (16)†
Continuous	–	349	.127	84.1	24.7 (77)
Persistence	–	358	.126	72.4	10.6 (500)
Climatology	.281	350	–	84.2	–

*Percent correct in this column refers to forecasts <1,000 ft for 5- and 11-hr forecasts and to forecasts <2,000 ft for 17-hr forecasts. The numbers in parentheses refer to the number of forecasts in this category.

†The best forecast system for each score and forecast projection.

3. The percent-correct score shows REEP to be superior to climatology for 5-hr forecasts but only slightly better for 11- and 17-hr forecasts. However, it should be noted that unconditional climatology always produces a forecast of category 5, which is certainly an undesirable characteristic, while the REEP system is capable of making correct forecasts for the lower categories. The Brier P-score and Allen utility score show definite improvement of REEP over climatology for all forecast projections.

Perhaps a better feeling for the forecast characteristics of the REEP system can be obtained from inspection of the contingency tables from which the verification scores were computed. The use of contingency tables in the verification required forecasts in categorical form. The REEP probability forecasts were transformed into categorical forecasts by three methods:

1. Maximizing the percent correct.
2. Maximizing the Allen utility score.
3. Maximizing the Heidke skill score.

The concept of maximizing these scores is discussed in the appendix. Tables 11, 12, and 13 show the contingency tables resulting from these three sets of forecasts (for the 5-hr forecast projection) paired with the respective observations.

Inspection of these tables shows the following:

1. The percent-correct method underforecasts all of the lower four categories by a considerable amount. Underforecasting (overforecasting) is defined as a situation where the total number of forecasts in a particular category is less than (greater than) the total number of observed in that category. For example, table 11 shows that the total number of forecasts in the category 0–1 is 44, while the total number observed in that category is 56. Therefore, the category 0–1 is underforecast.

2. The skill score method also underforecasts categories 1, 2, and 4 but not category 3. All of the lower four categories are forecast more often with the skill score method than with the percent-correct method.

TABLE 11.—*Verification of objective 5-hr ceiling height forecasts (REEP system) transformed from probability forecasts to maximize the number (or percent) correct. Ceiling categories are indicated in 100s of feet.*

Observed	Forecast					
	0–1	2–4	5–9	10–19	≥20	Total
0–1	18	7	4	2	25	56
2–4	15	33	14	6	53	121
5–9	4	15	53	23	88	183
10–19	0	11	25	22	212	270
≥20	7	12	12	20	1, 838	1, 889
Total	44	78	108	73	2, 216	2, 519

TABLE 12.—*Verification of objective 5-hr ceiling height forecasts (REEP system) transformed from probability forecasts to maximize the Allen score. Ceiling categories are indicated in 100s of feet.*

Observed	Forecast					
	0–1	2–4	5–9	10–19	≥20	Total
0–1	22	12	9	0	13	56
2–4	20	42	40	2	17	121
5–9	4	32	120	13	14	183
10–19	3	22	138	53	54	270
≥20	23	27	187	202	1, 450	1, 889
Total	72	135	494	270	1, 548	2, 519

TABLE 13.—*Verification of objective 5-hr ceiling height forecasts (REEP system) transformed from probability forecasts to maximize the skill score. Ceiling categories are indicated in 100s of feet*

Observed	Forecast					
	0–1	2–4	5–9	10–19	≥20	Total
0–1	18	10	6	2	20	56
2–4	16	35	28	13	29	121
5–9	4	21	80	33	45	183
10–19	1	18	60	65	126	270
≥20	7	19	46	80	1, 737	1, 889
Total	46	103	220	193	1, 957	2, 519

3. The Allen method overforecasts the lower three categories and is correct on the fourth category.

After forming the three contingency tables, we verified each set of forecasts by each of the three methods: percent correct, Allen utility score, and Heidke skill score. These scores are shown in table 14 along with the scores for persistence. The objective forecasts were better than persistence for all three scores but only when the categorical forecasts were made in such a way as to maximize that score. In other words, if categorical forecasts had been made with the objective of getting a large percent correct, and the forecasts were evaluated according to the skill score or Allen score, persistence would have been better than the objective forecasts.

TABLE 14.—*Three sets of tacegorical forecasts made from probability forecasts (REEP system) to maximize percent correct, Allen utility score, and Heidke skill score, respectively, verified according to each of the three scores*

Verification score	Objective forecasts score maximized			Persistence
	Percent correct	Allen utility score	Heidke skill score	
Percent correct	78. 0	66. 9	76. 8	77. 6
Allen utility score	437	521	458	476
Heidke skill score	0. 337	0. 353	0. 421	0. 416

6. FUTURE WORK

The single-station forecast system developed with the REEP technique (Allen 1970*a*) can be described as a sophisticated, local, conditional climatology. These equations use observed surface variables at the predictand station alone to produce probability forecasts for five categories of ceiling and visibility from 3 to 12 hr in advance. It should be possible to improve upon the REEP forecast system developed by the MOS approach by incorporating forecasts from the single-station system as predictors. In this combined system, the single-station predictors would provide information on local weather effects while the numerical model output would anticipate weather changes on the synoptic scale. The combined system should produce better results than either system used alone; this hypothesis will be tested as part of the TDL terminal weather prediction project.

The prediction of cloud bases, fog, wind, ceiling, and visibility by the MOS approach should improve as numerical prediction models improve and include more details of the lower atmosphere. Perhaps major improvement in aviation forecasting will come with the implementation of boundary-layer, numerical prediction models.

7. CONCLUSIONS

The use of the MOS approach in the development of ceiling height forecast equations has been discussed. Two forecast systems were developed: the REEP system, which produces probability forecasts of operationally significant ceiling height categories, and the continuous system, which produces specific ceiling height forecasts. In the continuous forecast system, we found that the use of the natural log transform $Y=\ln(X+0.001)$ gave the best results (in terms of percent correct) for the lower, more operationally significant ceiling categories when compared to other transform functions.

Independent data verification indicated the following:

1. The REEP equations should contain 14 to 20 predictors, and the continuous equations should include about 10 predictors.

2. The use of SAM and PE forecast output in the REEP equations improved upon results obtained from using 0700 GMT initial data alone.

3. The REEP system was generally better than the continuous system, persistence, and climatology.

Therefore, the MOS approach through REEP equations shows definite promise for operational use.

APPENDIX: DEFINITION OF VERIFICATION SCORES

This study describes the development of two ceiling height forecast systems. The first system, called REEP, produces probability forecasts; and the second system, called continuous, produces specific ceiling height forecasts. The verification of these two systems on independent data involves the computation of the Brier P-score, Allen utility score, percent correct, and Heidke skill score. This appendix defines these scores and discusses the transformation of probability forecasts into categorical forecasts to maximize a particular score.

Allen (1970*b*) states ". . . It is desirable that statements of the probability of a weather event be *reliable;* that is, over a period of time the event should actually occur with the frequency implied by the probability forecast. It is also desirable that the probabilities be as close to zero or to 100 percent as possible when the event does not occur or does occur, respectively." The Brier P-score, P, (Brier 1950) measures these two characteristics probability forecasts and is given by

$$P=\frac{1}{N}\sum_{j=1}^{r}\sum_{i=1}^{N}(f_{ij}-E_{ij})^2 \tag{2}$$

where on each of N occasions an event can happen in only one of r possible classes, and $f_{i1}, f_{i2}, \ldots, f_{ir}$ represent the forecast probabilities that the event will occur in classes 1, 2, . . ., r, respectively. If the r classes are chosen to be mutually exclusive and exhaustive,

$$\sum_{j=1}^{r} f_{ij}=1 \tag{3}$$

for each and every i=1, 2, . . . , N. E_{ij} takes the value 1 or 0 according, respectively, to whether the event occurred in class j or not. For perfect forecasting, the Brier P-score will have a value of zero and for the worst possible forecasting a value of two.

The percent correct, Heidke skill score, and Allen utility score were computed from contingency tables. A typical contingency table has the form shown in table 15. For instance, for a particular case, if both the forecast and observed ceiling fall into category 1, then the value of A is increased by one. If the forecast value falls into category 1 and the observed value into category 2, then B is increased by one, and so forth.

The percent correct, PC, of the total number of forecasts, T, is computed by

$$PC=\frac{R}{T}\times 100 \tag{4}$$

TABLE 15.—*A typical three-category contingency table*

Observed category	Forecast category			
	1	2	3	Total
1	A	E	I	M
2	B	F	J	N
3	C	G	K	O
Total	D	H	L	T

TABLE 16.—*The Allen utility matrix used to judge the usefulness of ceiling and visibility forecasts*

Observed category	Forecast category				
	1	2	3	4	5
1	1.0	0.6	0.1	0.0	0.0
2	.7	.9	.4	.05	.0
3	.2	.5	.7	.2	.0
4	.0	.1	.3	.45	.1
5	.0	.0	.05	.1	.15

where the number of correct forecasts, R, is given by

$$R=A+F+K. \tag{5}$$

The Heidke skill score, SS, is computed by

$$SS=\frac{R-P}{T-P} \tag{6}$$

where P is the number of forecasts expected to be correct by chance. (See, e.g., Panofsky and Brier 1958.)

The computation of the Allen utility score involves the use of a utility matrix. A utility matrix essentially shows a series of weighting factors that are meant to represent the usefulness or utility of forecasts to the user. This concept is discussed in detail by Glahn (1964).

The utility matrix used in this study and shown in table 16 was devised by Allen after consultation with forecasters at several aviation forecast centers (Glahn 1964). This matrix may not be too different from that of an actual utility matrix of an airline, and it has been used by Enger et al. (1962) in evaluating ceiling height forecasts at seven terminals.

Inspection of table 16 shows that a correct forecast of category 1 receives a weight of 1.0, while a correct forecast of category 5 is given a weight of only 0.15. Also, a near miss such as a forecast of category 1 when category 2 is observed is weighted by 0.7. The Allen utility score (AUS) is given by

$$\text{AUS}=\sum_{j=1}^{5}\sum_{i=1}^{5} W_{ij}Z_{ij} \tag{7}$$

where W represents the weights shown in table 16, and Z represents the values in the corresponding boxes of a 5 by 5 contingency table. This score, therefore, has the following characteristics: (1) more credit is given for correct forecasts of the lower, more operationally significant, ceiling categories than for correct forecasts in higher categories; and (2) credit is given for near misses.

Since the computation of the Allen utility score, percent correct, and skill score involves the use of contingency tables, it was necessary to convert the probability forecasts produced by the REEP system into categorical forecasts. That is, for a given event, the REEP system produces five probability forecasts, $f_1, f_2, \ldots, f_5$, corresponding to the five ceiling height predictand categories (table 2). Given the five probabilities, the question is, "Which ceiling category should be forecast?" The answer depends on which verification score is being used, because it is theoretically possible (provided the probability forecasts are unbiased) to maximize a particular score by picking the category according to a certain rule. One of the important features of probability forecasts is that they lend themselves to such treatment.

For instance, the Allen utility score is maximized as follows. Let f_1, f_2, f_3, f_4, and f_5 be the forecast probabilities for a particular event. Five quantities are then computed:

$$E_1=\sum_{j=1}^{5} W_{j1}f_j,\ E_2=\sum_{j=1}^{5} W_{j2}f_j,\ \ldots,\ E_5=\sum_{j=1}^{5} W_{j5}f_j \quad (8)$$

where the W values are obtained from the Allen matrix. For example, W_{j1}, where j goes from one to five, represents the five weights given in the forecast category 1 column of the scoring matrix shown in table 16. The maximum E indicates the proper categorical forecast. Thus, if E_2 is the largest, ceiling category 2 is forecast.

The percent-correct score can be maximized in a similar manner except that the scoring matrix would look like that shown in table 17. Note that a correct forecast of category 1 is given the same weight as a correct forecast of category 5. Also, no credit is given for near misses. The use of this scoring matrix and eq (8) would give the same result as forecasting the category having the highest probability.

Bryan and Enger (1967) developed methods for maximizing various skill scores, given probability forecasts. Their method for maximizing the Heidke skill score [defined in eq (6)] was used in this study. No details will be given here, but briefly the decision rule is to choose ceiling category k instead of j on trail t' if

$$f_k(t')-f_j(t')>(1-h)(p_k-p_j) \quad (9)$$

where f is defined as in eq (8), p_k and p_j are the climatological probabilities of the ceiling categories k and j, and h is the optimum skill score computed on the dependent data sample. As recommended by Bryan and Enger (1967), the optimum skill score on the dependent data sample was determined by starting with a value of h given by persistence and then using eq (9) in an iterative procedure to determine the final, optimum skill score for each verification station. (Note that data from different

TABLE 17.—*Scoring matrix for maximizing the percent-correct score*

Observed category	Forecast category 1	2	3	4	5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

stations were *not* combined for this purpose.) This optimum skill score was then used in applying eq (9) to the independent data sample.

ACKNOWLEDGMENTS

This work was supported by the Federal Aviation Administration under Interagency Agreement FA67WAI-131. We wish to thank the many members of TDL who contributed to this study through computer programming and data preparation in the SAM project. We also thank Pat Robinson for typing the manuscript.

REFERENCES

Allen Roger A., "Operational Evaluation of a Ceiling and Visibility Prediction Technique," *Report* No. FAA-RD-70-17, Federal Aviation Administration, Washington, D.C., Dec. **1969**, 19 pp.

Allen, Roger A., "Single Station Prediction of Ceiling and Visibility," *Report* No. FAA-RD-70-26, Federal Aviation Administration, Washington, D.C., Apr. **1970***a*, 48 pp.

Allen Roger A., "Predicting Ceiling and Visibility With Boolean Predictors," *Report* No. FAA-RD-70-37, Federal Aviation Administration, Washington, D.C. May **1970***b*, 16 pp.

Annett, John R., Glahn, Harry R., and Lowry, Dale A., "The Use of Model Output Statistics (MOS) to Estimate Daily Maximum Temperatures," *NOAA Technical Memorandum* NWS TDL-45, U.S. Department of Commerce, Techniques Development Laboratory, Silver Spring, Md., Mar. **1972**, 14 pp.

Barrientos, Celso S., "An Objective Method for Forecasting Winds Over Lake Erie and Lake Ontario," *ESSA Technical Memorandum* WBTM TDL 34, U.S. Department of Commerce, Techniques Development Laboratory, Silver Spring, Md., Aug. **1970**, 20 pp.

Brier, Glenn W., "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, Vol. 78, No. 1, Jan. **1950**, pp. 1-3.

Bryan, Joseph G., and Enger, Isadore, "Use of Probability Forecasts to Maximize Various Skill Scores," *Journal of Applied Meteorology*, Vol. 6, No. 5, Oct. **1967**, pp. 762-769.

Enger, Isadore, Reed, L. J., and MacMonegle, J. E., "An Evaluation of 2-7-Hr Aviation Terminal-Forecasting Techniques," *Technical Publication* 20, Contract No. FAA-BRD-363, The Travelers Research Center, Inc., Hartford, Conn., Oct. **1962**, 36 pp.

Enger, Isadore, Russo, John A., Jr., and Sorenson, Edna L., "A Statistical Approach to 2-7-Hour Prediction of Ceiling and Visibility," *Technical Report* No. 2, Contract No. Cwb-10704, The Travelers Research Center, Inc., Hartford, Conn., Sept. **1964**, 48 pp.

Glahn, Harry R., "The Use of Decision Theory in Meteorology," *Monthly Weather Review*, Vol. 92, No. 9, Sept. **1964**, pp. 383-388.

Glahn, Harry R., "A Method for Predicting Surface Winds," *ESSA Technical Memorandum* WBTM TDL 29, U.S. Department of Commerce, Techniques Development Laboratory, Silver Spring, Md., Mar. **1970**, 18 pp.

Glahn, Harry R., and Lowry, Dale A., "An Operational Method for Objectively Forecasting Probability of Precipitation," *ESSA Technical Memorandum* WBTM TDL 27, U.S. Department of Commerce, Techniques Development Laboratory, Silver Spring, Md., Oct. **1969**, 24 pp.

Glahn, Harry R., Lowry, Dale A., and Hollenbaugh, George W., "An Operational Subsynoptic Advection Model," *ESSA Technical Memorandum* WBTM TDL 23, U.S. Department of Commerce, Techniques Development Laboratory, Silver Spring, Md., July **1969**, 26 pp.

Glahn, Harry R., and Lowry, Dale A., "The Use of Model Output Statistics in Objective Weather Forecasting," *Journal of Applied Meteorology*, Vol. 11, No. 8, Dec. **1972**, pp. 1203–1211.

Harris, R.G., Bryan, J. G., and MacMonegle, J. E.," Terminal Weather Prediction Studies," *Technical Note* 3, Contract No. AD19(626)–16, The Travelers Research Center, Inc., Hartford, Conn., May **1963**, 103 pp.

Keaty, John, and Allen, Roger A., "Development of Base Techniques for Ceiling and Visibility Prediction," *Report* No. FAA–RD–70–22, Federal Aviation Administration, Washington, D.C., Feb. **1970**, 21 pp.

Klein, William H., "Application of Synoptic Climatology and Short-Range Numerical Prediction to Five-Day Forecasting," *Research Paper* No. 46, U.S. Department of Commerce, Weather Bureau, Washington, D.C., June **1965**, 109 pp.

Miller, Robert G., "A Statistical Procedure for Screening Predictors in Multiple Regression," *Final Report*, Contract No. AF19(604)–1590, The Travelers Weather Research Center, Inc., Hartford, Conn., Dec. 31, **1958**, 238 pp. (see pp. 86–95).

Miller, Robert G., "Regression Estimation of Event Probabilities," *Technical Report* No. 1, Contract No. 1 Cwb–10704, The Travelers Research Center, Inc., Hartford, Conn., Apr. **1964**, 153 pp.

Panofsky, Hans A. and Brier, Glenn W., *Some Applications of Statistics to Meteorology*, College of Mineral Industries, The Pennsylvania State University, University Park, **1958**, 224 pp.

Roberts, C. F., "Terminal Weather Forecasting-Current Programs and Prospects for the Future," *Proceedings of the 1967 Annual Assembly Meeting of the Radio Technical Commission for Aeronautics, September 19–20, 1967, Washington, D.C.*, Radio Technical Commission for Aeronautics, Washington, D.C., **1967**.

Russo, John A., Enger, Isadore, and Merriman, Guy T., "A Statistical Approach to the 12–48-Hr Prediction of Precipitation Probability," *Final Report*, Contract No. Cwb–11100, The Travelers Research Center, Inc., Hartford, Conn., Aug. **1966**, 107 pp.

Shuman, Frederick G., and Hovermale, John B., "An Operational Six-Layer Primitive Equation Model," *Journal of Applied Meteorology*, Vol. 7, No. 4, Aug. **1968**, pp. 525–547.

Younkin, Russell J., LaRue, Jerrold A., and Sanders, Frederick, "The Objective Prediction of Clouds and Precipitation Using Vertically Integrated Moisture and Adiabatic Motions," *Journal of Applied Meteorology*, Vol. 4, No. 1, Feb. **1965**, pp. 3–17.

[Received March 29, 1972; revised May 30, 1972]

UDC 551.515.43'.83:551.507.362.2:551.576.1(084.1)(759)"1972.02.17"

PICTURE OF THE MONTH
Thin Line Convection

FRANCES C. PARMENTER and STANLEY W. WRIGHT [1]—*Applications Group, National Environmental Satellite Service, NOAA, Suitland, Md.*

Satellite photographs often reveal a bright rope of convection along the leading edge of a frontal band. These thunderstorms are usually not severe and tops range from 18,000 to 30,000 ft. (anonymous 1967). They are seen over both land and water areas during the colder season of the year. In this case, both the ATS 3 satellite and Miami, Fla., radar recorded the passage of such a convective line.

On Feb. 17, 1972, a cold front was moving across the Florida Peninsula. The main Low was off the North Carolina coast with the front extending from north of Palm Beach, Fla., southwestward into the Gulf of Mexico (fig. 1). During the early morning, only light winds, showers or light rain, and fog accompanied its passage. Another trough was drawn from the North Carolina border through central Georgia and into the gulf.

The early morning (1328 GMT) satellite view (fig. 2) shows the clouds associated with these systems. A low winter sun angle causes the higher clouds to be highlighted along their eastern edges and to cast large shadows along their western edge. Here, the brightest, most convective clouds (J, K) can be seen on the leading edge of the front (L–M). This frontal zone cloudiness extends back to Alabama and Georgia.

In the next view (fig. 3), taken at 1449 GMT, the line of convection (J) has made some southward progress over western Florida; while the eastern edge has remained almost stationary. Surface reports at 1500 GMT show 15-kt northwesterly winds at Key West, Fla., while Miami and Palm Beach, where the line is stationary, report southwesterly winds at 15 kt. Figure 4 shows the radar echoes associated with this line of convection at 1440 GMT. Note the slight bulge (P) in the forward-moving segment of the line. The northern portion of the line (K, fig. 3)

[1] University of St. Thomas, Houston, Tex., student temporarily assigned to the Applications Group, NESS.

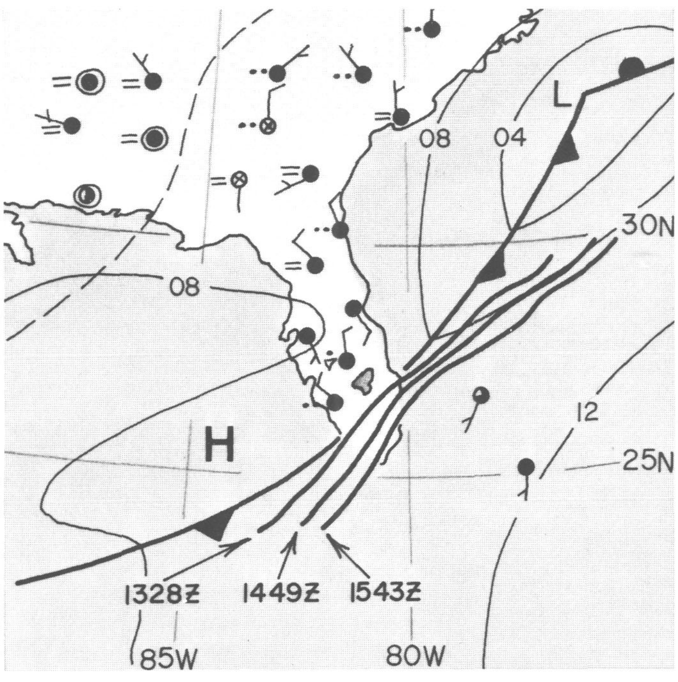

FIGURE 1.—The 1200 GMT surface map and positions of the thin line of convection.

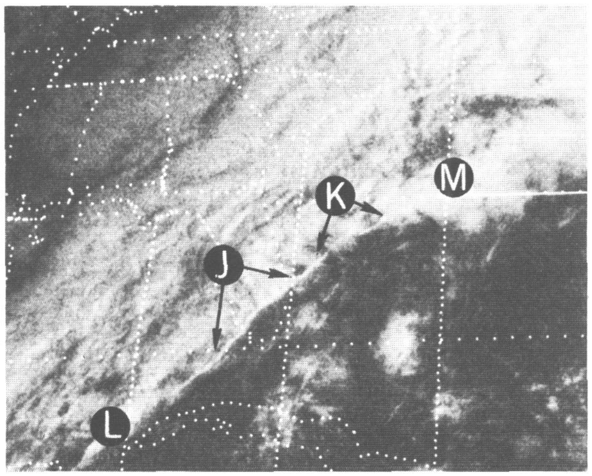

FIGURE 2.—ATS 3 satellite photograph, 1328 GMT, Feb. 17, 1972.

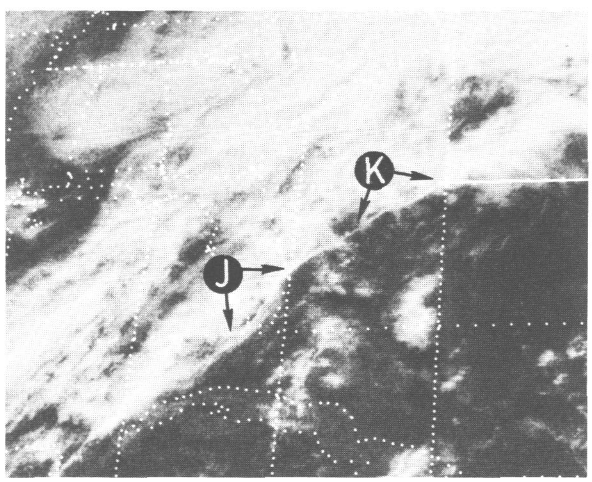

FIGURE 3.—ATS 3 satellite photograph, 1449 GMT, Feb. 17, 1972.

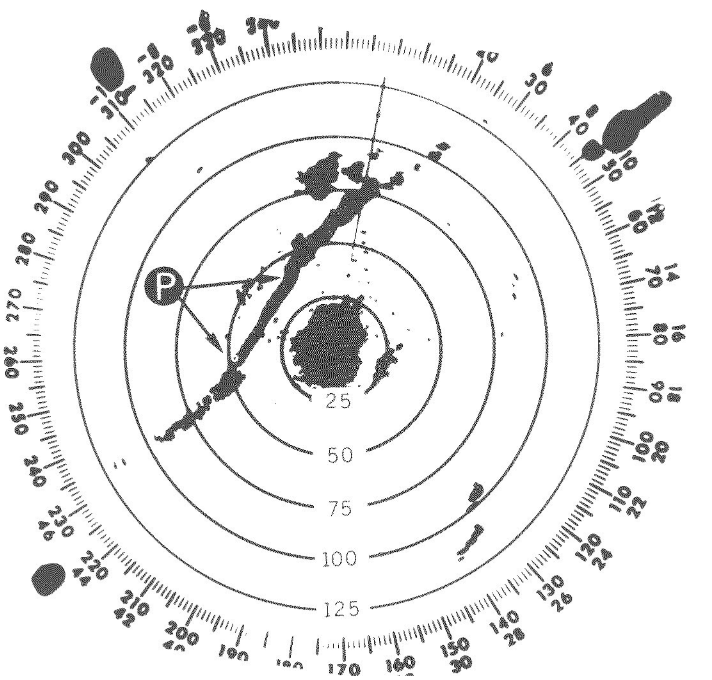

FIGURE 4.—Photograph of Miami, Fla., radarscope, 1440 GMT, Feb. 17, 1972.

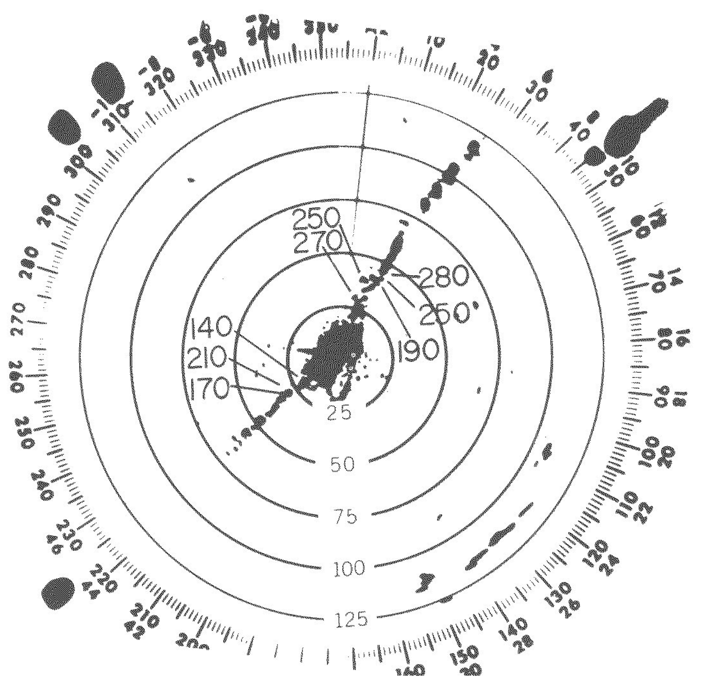

FIGURE 6.—Photograph of Miami, Fla., radarscope, 1555 GMT, Feb. 17, 1972.

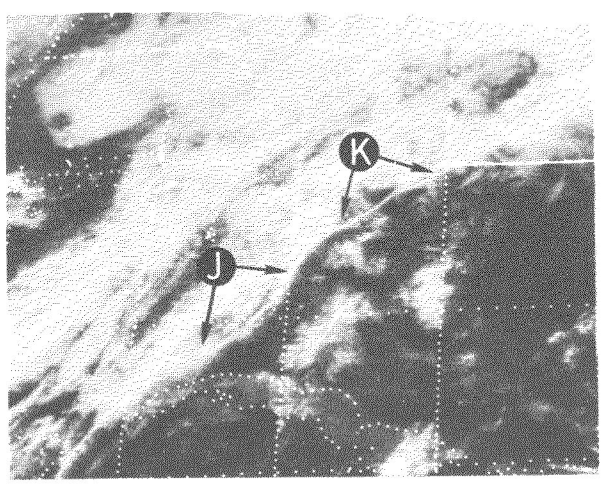

FIGURE 5.—ATS 3 satellite photograph, 1543 GMT, Feb. 17, 1972.

appears to have moved out from under the more solid clouds to become the leading edge.

In the last satellite view (fig. 5), taken at 1543 GMT, part of the line (J) was located through Miami and the Florida Keys. Figure 6 shows the radar observations at 1555 GMT. Vertical measurements of some of these echoes (noted on fig. 6) showed tops ranging from 14,000 to 28,000 ft. In the 75 min between observations, this line of echoes moved 25 n.mi. to the southeast. This movement is slightly faster than the surface wind speed reported at 1500 GMT and has a component opposite in direction to the southwesterly winds aloft throughout this period. The increased southward speed could be due to the downrush of the accompanying, precipitation-cooled air. During the summer months, we have observed a similar phenomenon accompanying thunderstorm clusters. These systems are much more active, but they often cause an arc of clouds to travel outward from the cluster at a speed much greater than the surface wind.

REFERENCE

Anonymous, "Picture of the Month," *Monthly Weather Review*, Vol. 95, No. 2, Feb. **1967**, p. 98.